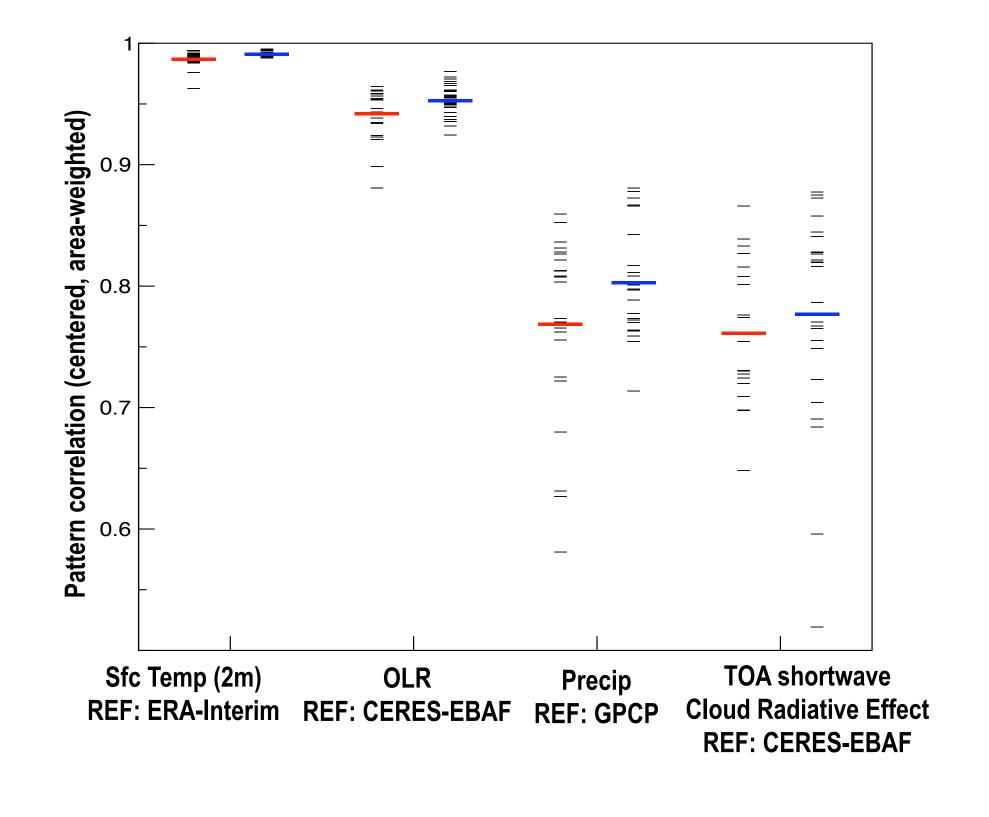
Climate Model Performance Metrics: Research and Community Leadership

A cross-cutting activity within LLNL's climate work: Using both well-established and research driven performance metrics to succinctly quantify simulation agreement with observations

CMIP model performance Are models improving?

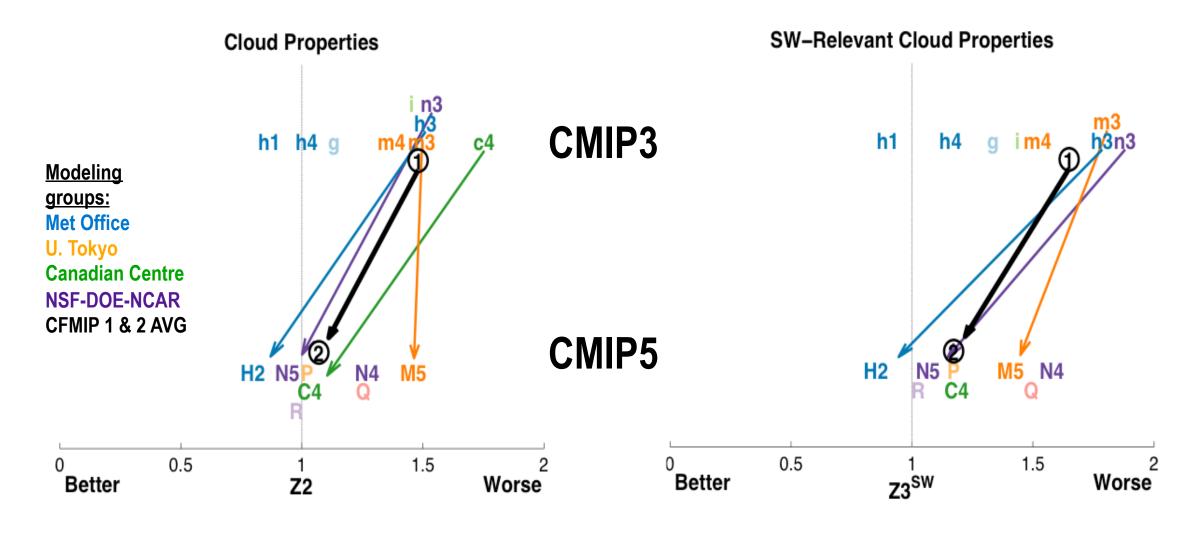
CMIP3 and CMIP5 vs. observations Annual and global mean pattern correlation



- Incremental improvement in the ensemble multi-model average across variables
- Improvements more apparent in the poorer performing models

Tracking changes in the simulation of clouds using the ISCCP satellite simulator: CMIP3 to CMIP5

Findings from Klein et al. (submitted, JGR-Atmospheres):



Cloud property scalar measure (Z2): quantifies fidelity in simulating cloud-top pressure and optical depth in different categories of optically intermediate and thick clouds at high, middle, and low-levels of the atmosphere

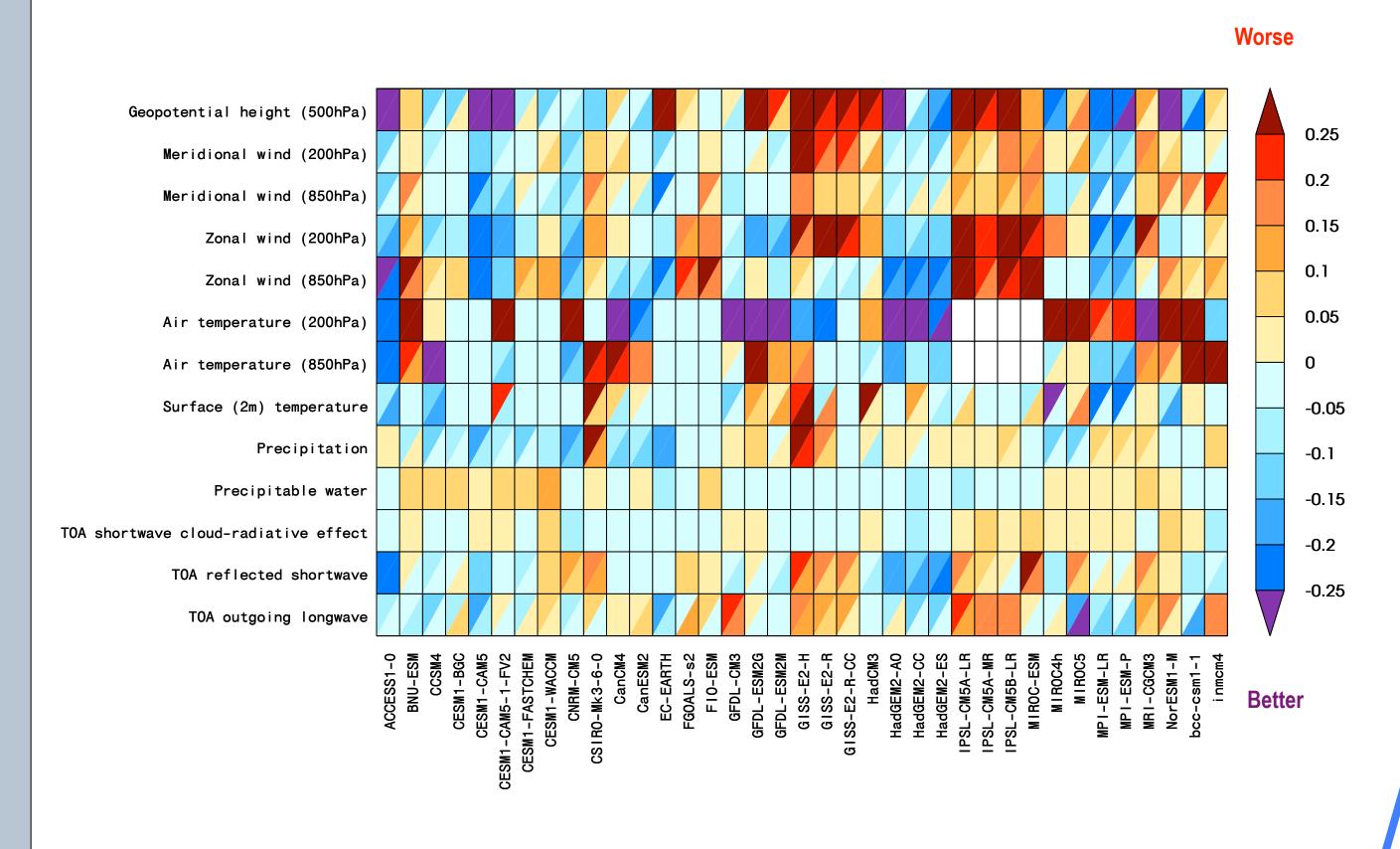
SW relevant cloud properties (**Z**3): measures the impacts on top-of-atmosphere shortwave radiation in the same categories measured by **Z**2.

- Demonstrable progress by four individual centers at simulating cloud properties, largely due to better simulation of the amounts of optically intermediate and thick clouds
- Better simulation of cloud properties has significantly reduced common compensating errors ("too few but too bright") that are frequently present with a TOA SW and LW balance

Additional performance metrics work is described in other LLNL SFA review posters (P. Durack, K. Sperber and D. Kim, Y. Zhang et al., S. Xie et al.) and presentations (B. Santer, K. Sperber)

CMIP model performance Relative skill measures

A portrait plot: CMIP5 annual cycle metrics for multiple fields

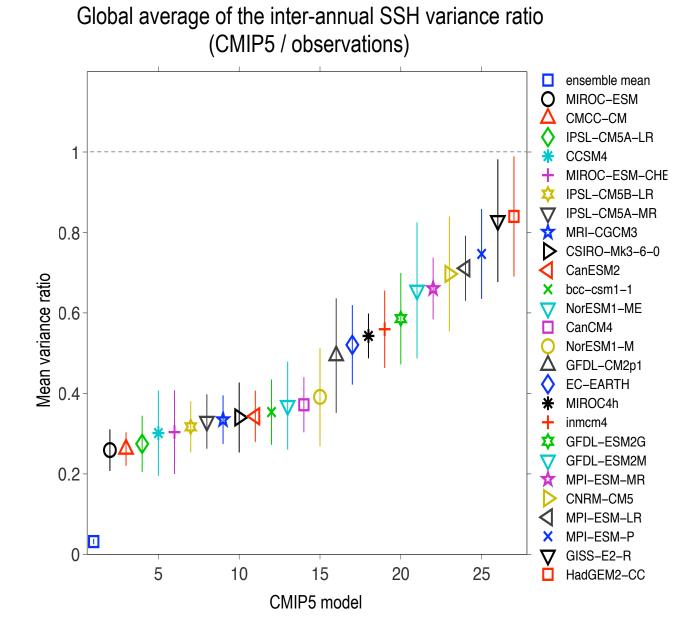


Global annual cycle climatology (1980-2005) computed from CMIP5 historical experiments. The error measure is a space-time root-mean square error (RMSE), which, treating each variable separately, is portrayed as a relative error by normalizing the result by the median error of all model results (Gleckler at al., 2008). For example, a value of 0.20 indicates that a model's RMSE is 20% larger than the typical CMIP5 error for that variable, whereas a value of -0.20 means the error is 20% smaller than the typical error. A diagonal splits each grid square, showing the relative error with respect to both the default (ERA-Interim and CERES-EBAF, upper left triangle) and the alternate (NCEP-NCAR reanalysis and ERBE, lower right triangle) reference data sets. Some model data is not yet available (white squares). These results are being routinely updated as the CMIP5 database continues to populate, and therefore are preliminary.

- Provides a quantitative synthesis of how some models clearly simulate the annual cycle more realistically than others
- For most variables the selection of reference data does not radically change estimate of relative model performance
- The results from these and other well-established performance metrics are to be made available to the research community via the WCRP metrics panel. The database of results, analysis code and observations will be provided to modeling groups, giving them the possibility to incorporate a multi-model analysis directly into their model development process

CMIP5 dynamic sea-surface height (SSH) compared to satellite altimeter data (1988-2005)

Findings from Landerer (JPL), Gleckler and Lee (submitted, JGR-Oceans):

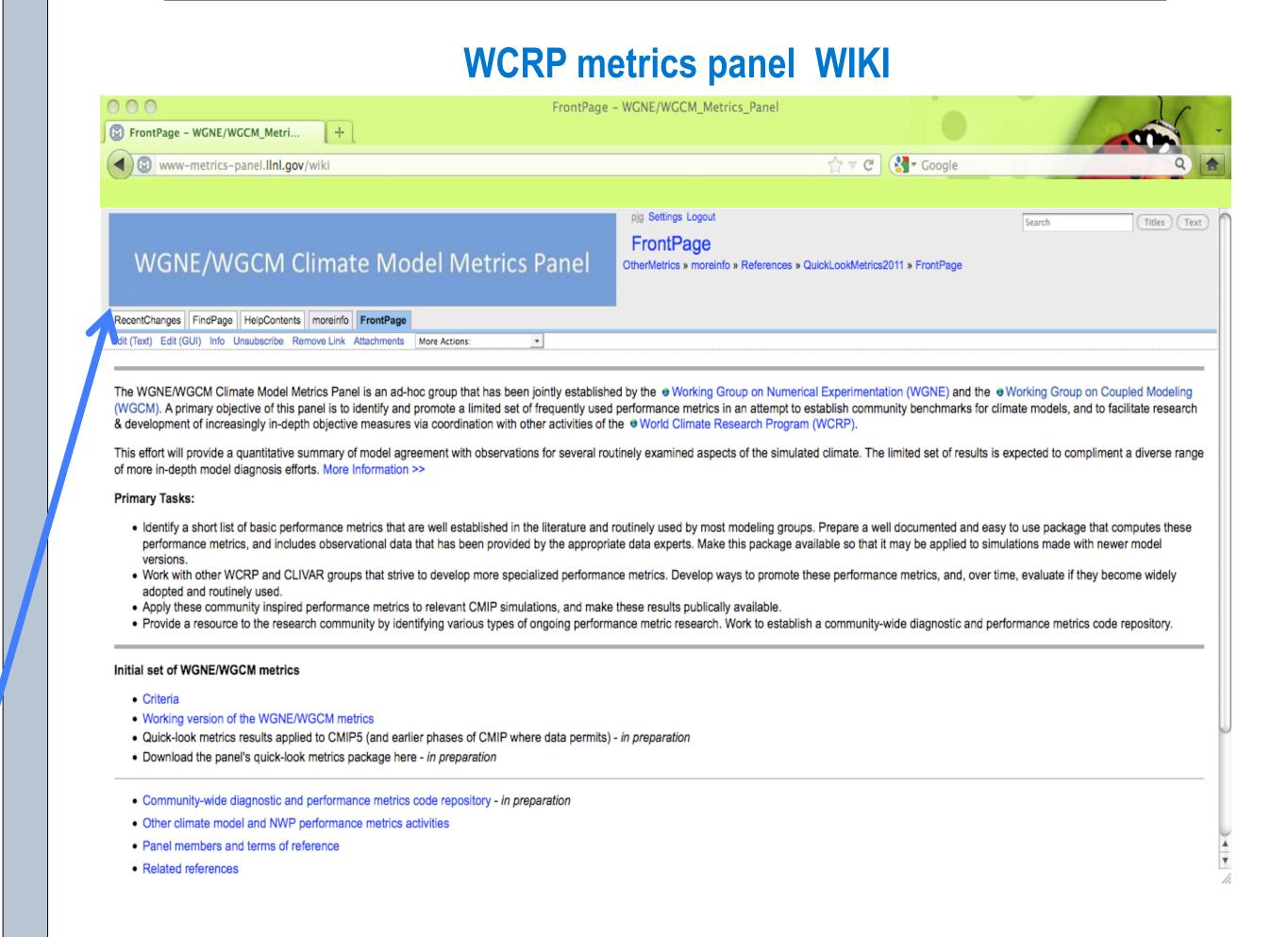


The symbols depict the 1988-2005 variance-ratio (model/obs) from observations (ALVISO) and available CMIP5 models. The vertical lines give the 1-sigma range of the globally average variance ratios from a sliding 18-year window analysis over the historical run of each model.

- Satellite altimetry provides high quality near global SSH measurements
- CMIP5 models underestimate inter-annual variance, with only a few models coming close to observations
- Systematic underestimate evident in the equatorial Central to East Pacific, western Pacific warm pool, tropical Indian Ocean, western boundary currents (not shown)
- Although a shorter record than in-situ based steric estimates, as independent measurements altimetry may help additional research into how well models simulate longer time scale variability

From benchmark experiments to community-based performance testing

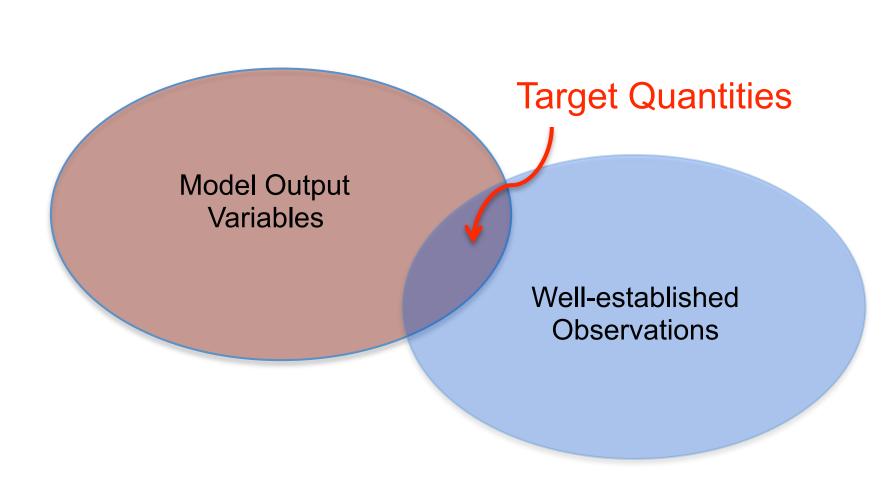
PCMDI is leading an international effort to establish routine tests of climate model performance. This activity is coordinated through the WCRP, and engages all leading modeling centers



Bridging the gap between climate models and observations: OBS4MIPs

In collaboration with JPL and over a dozen NASA mission science teams, DOE/PCMDI has been working to make selected satellite products more widely accessible for CMIP5 research. This pilot project is called Obs4MIPs. Criteria for a product to be made available via Obs4MIPs include:

- The observations must be directly comparable to a model output field defined as part of CMIP5
- Include documentation with information of relevance to model analysts
- Traceable product version changes (critical for the metrics panel effort)
- Organized analogous to CMIP5 (format, data structure, etc.)
- Served through ESGF



- To date, most observations available via Obs4MIPs have been contributed by NASA, but DOE's Atmospheric Radiation Measurement Best Estimate (ARMBE) is also included.
- With participation in the WCRP's new Data Advisory Council (WDAC), PCMDI is now working to expand this effort to enable other data providers. NOAA and the European Space Agency, e.g., have expressed their intent to contribute.